

**AFRL-RI-RS-TR-2009-31**  
**Final Technical Report**  
**February 2009**



# **A TWO-DIMENSIONAL DEOXYRIBONUCLEIC ACID (DNA) MATRIX BASED BIOMOLECULAR COMPUTING AND MEMORY ARCHITECTURE**

The Research Foundation of the State University of New York

*APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.*

STINFO COPY

**AIR FORCE RESEARCH LABORATORY  
INFORMATION DIRECTORATE  
ROME RESEARCH SITE  
ROME, NEW YORK**

## NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the 88<sup>th</sup> ABW, Wright-Patterson AFB Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (<http://www.dtic.mil>).

AFRL-RI-RS-TR-2009-31 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE DIRECTOR:

/s/

THOMAS E. RENZ  
Work Unit Manager

/s/

EDWARD JONES, Deputy Chief  
Advanced Computing Division  
Information Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

**REPORT DOCUMENTATION PAGE***Form Approved*  
**OMB No. 0704-0188**

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Service, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.

**PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.****1. REPORT DATE (DD-MM-YYYY)**

FEB 09

**2. REPORT TYPE**

Final

**3. DATES COVERED (From - To)**

Nov 05 – Oct 08

**4. TITLE AND SUBTITLE**A TWO-DIMENSIONAL DEOXYRIBONUCLEIC ACID (DNA) MATRIX  
BASED BIOMOLECULAR COMPUTING AND MEMORY  
ARCHITECTURE**5a. CONTRACT NUMBER**

FA8750-06-2-0002

**5b. GRANT NUMBER**

N/A

**5c. PROGRAM ELEMENT NUMBER**

61101E

**6. AUTHOR(S)**

Anthony Macula, Russell Deaton, and Junghuei Chen

**5d. PROJECT NUMBER**

459T

**5e. TASK NUMBER**

DN

**5f. WORK UNIT NUMBER**

AM

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**The Research Foundation of the State University of New York  
35 State Street  
Albany, NY 12207-2826**8. PERFORMING ORGANIZATION  
REPORT NUMBER**

N/A

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**AFRL/RITC  
525 Brooks Rd.  
Rome NY 13441-4505**10. SPONSOR/MONITOR'S ACRONYM(S)**

N/A

**11. SPONSORING/MONITORING  
AGENCY REPORT NUMBER**  
AFRL-RI-RS-TR-2009-31**12. DISTRIBUTION AVAILABILITY STATEMENT**

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED. PA# 2009-0331

DATE CLEARED: 30-JAN-09

**13. SUPPLEMENTARY NOTES****14. ABSTRACT**

The primary goal of this research was the development of an enabling technology for DNA computing. It focused on the development of two distinct DNA-based nanotechnologies, aqueous computing and DNA arrays. Once merged, these DNA-based nanotechnologies serve as a platform for a hybrid silicon-DNA computing architecture. The aqueous side is where the computing happens and the array side is where the results are preprocessed. The array information passes to an electronic computer for signal processing type decoding. The intent was to develop a computing basis to overcome the exponential time complexity of many discrete mathematical problems so that they can be solved in linear real time. This research addresses needs of the Air Force because computationally hard (NP) problems are critical to logistics, scheduling and national security. Moreover, the DNA nanotechnologies developed in this research has the potential to enhance Biotagants which provide environmental awareness in the field.

**15. SUBJECT TERMS**

DNA Computing, Molecular Computing, Aqueous Computing, Hybrid Computing

**16. SECURITY CLASSIFICATION OF:**a. REPORT  
Ub. ABSTRACT  
Uc. THIS PAGE  
U**17. LIMITATION OF  
ABSTRACT**

UU

**18. NUMBER  
OF PAGES**

25

**19a. NAME OF RESPONSIBLE PERSON**

Thomas E. Renz

**19b. TELEPHONE NUMBER (Include area code)**

N/A

## TABLE OF CONTENTS

<b>1.0</b>	<b>SUMMARY</b>	<b>1</b>
<b>2.0</b>	<b>INTRODUCTION</b>	<b>2</b>
	<b>2.1</b> <b>Modified Nucleotides</b>	<b>2</b>
	<b>2.2</b> <b>DNA Arrays</b>	<b>2</b>
<b>3.0</b>	<b>METHODS, ASSUMPTIONS, PROCEDURES</b>	<b>5</b>
	<b>3.1</b> <b>Encoding Strategy</b>	<b>5</b>
	<b>3.2</b> <b>Coupling the Aqueous Method and SAT</b>	<b>6</b>
	<b>3.3</b> <b>Construction of High Density 2-D Array</b>	<b>7</b>
<b>4.0</b>	<b>RESULTS, DISCUSSION</b>	<b>10</b>
	<b>4.1</b> <b>Mathematical Computation with Labeled Nucleotides</b>	<b>10</b>
	<b>4.2</b> <b>Using SynDCode to Test 2-D Array</b>	<b>12</b>
	<b>4.3</b> <b>Obtaining a Dense 2-D Array</b>	<b>15</b>
<b>5.0</b>	<b>CONCLUSIONS</b>	<b>18</b>
	<b>5.1</b> <b>Other Applications</b>	<b>19</b>
<b>6.0</b>	<b>REFERENCES</b>	<b>20</b>

## LIST OF FIGURES

<b>1</b>	<b>Encoding Strategy</b>	<b>5</b>
<b>2</b>	<b>Signature Comparison for SAT Using Multiple 2-D Array</b>	<b>7</b>
<b>3</b>	<b>Separation of DNA Code on 2-D array</b>	<b>9</b>
<b>4</b>	<b>Steps in Computing p OR q</b>	<b>10</b>
<b>5</b>	<b>Dot Blots Results</b>	<b>13</b>
<b>6</b>	<b>Verification of Temperature Dimension</b>	<b>11</b>
<b>7</b>	<b>SynDCode Test Strands</b>	<b>14</b>
<b>8</b>	<b>Large Code Dispersion</b>	<b>15</b>
<b>9</b>	<b>Sufficient Strand Length for Error -Free Codes of Given Sizes</b>	<b>15</b>
<b>10</b>	<b>Making and Using a Dense Array</b>	<b>16</b>
<b>11</b>	<b>Verification of Dense Array Construction</b>	<b>17</b>

## **1.0 SUMMARY**

In this project two parallel tasks are merged in to a single deoxyribonucleic acid, DNA, computing architecture. A new approach to aqueous computing, where modified nucleotides are used to 'write' on double-stranded DNA molecules and a novel '2-D' method for DNA array construction are developed. The aqueous side is where the computing happens and the array side is where the results are decoded.

On the aqueous side, binding proteins specific to each modification are used to selectively isolate DNA fragments and this selectivity is used to do computation. As there are millions of molecules with corresponding binding proteins, this approach has the potential to yield unlimited computing power as compared with other aqueous computing methods.

On the array side, current genomic DNA array fabrication methods are impractical for information processing applications because they are expensive, difficult to produce for universal applications and lack a storage capacity that can rival silicon based systems. However, all DNA arrays have parallel nucleic acid detection capabilities and this model can be used to support DNA based information processing. To go beyond the current DNA array fabrication methods, a new method of DNA array construction that uses both the DNA melting temperature and DNA secondary structure dimensions was developed.

## **2.0 INTRODUCTION**

### **2.1 Modified Nucleotides**

The successful use of molecular biology methods as computational tools by Adelman [1] has led to the development of a variety of DNA computing techniques. Aqueous computing as proposed by Head and Gal [2] uses large numbers of initially identical molecules, such as DNA, which serve as memory registers in a fluid environment. Bit values (e.g., 0, 1) can be “written” on the molecule and subsequently “read” to determine solutions to computational problems. A major advantage of aqueous computing is that the fluid memory can be proportioned out and mixed back together so that problems requiring an exponential number of steps to solve conventionally, involve only a linear number of steps in this method.

Here a new approach to aqueous computing where the “writing” step involves the labeling of DNA with different molecules incorporated into nucleotides is investigated. These modified nucleotides are used to selectively isolate DNA fragments by binding proteins specific to the label of interest. An implementation strategy for using these binding proteins in aqueous computing is given with the successful incorporation of 4 different modified nucleotides into DNA and isolation of DNA labeled with two of these modified nucleotides using binding proteins.. The commercial availability of modified nucleotides and corresponding binding proteins allows for larger computations while the ease of incorporation speeds up computation time.

### **2.2 DNA Arrays**

DNA microarrays have sizable parallel nucleic acid detection capabilities and this model can be used to support DNA based information processing [3], [4]. However, current arrays are expensive and don't have an information capacity that can rival silicon based systems. Here a novel method of DNA array construction that uses both the melting temperature and the secondary structure dimensions is explored.

This method of construction consists of separating up to  $\sim 10^{17}$  different 50-mer strands in a 2-D matrix that, by generating a hybridization signature, is capable of recalling or monitoring any input DNA. Thus the 2-D array can be thought of as a DNA memory. Moreover, because of the ability to profile synthetic or genomic DNA, the 2-D matrix has numerous potential applications beyond information processing in the fields of bio-defense monitoring and medical diagnostics.

Current photolithographic or ink-jet technologies for the manufacture of DNA arrays have common drawbacks. One major impediment for information processing applications is that DNA sequences in the array must be chosen and *isolated* in advance of manufacture, thus even under automation, constructing an array with more than  $10^6$  memory locations is impractical. Therefore, using current microarray technologies, it seems impossible to build universal microarray with an information density that can rival silicon. The 2-D universal DNA matrix model attempts to address this problem. It is a membrane with large number of oligonucleotides *that become immobilized on the surface in a self assembled process that is reproducible*. It is constructed by a method by which DNA sequences self-separate on the surface directed by their physical properties [5]-[7]. Such a 2-D matrix is capable of monitoring or profiling the existence of any DNA input by outputting a unique hybridization signature that corresponds to unique pattern that can be mapped to the unique locations of the probes on the membrane. The fact that  $10^{17}$  strands can be uniquely separated can be exploited to solve complex mathematical problems like SAT [1].

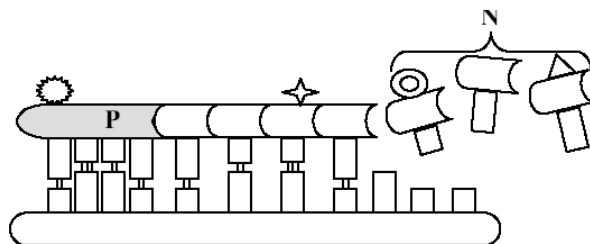


Due to the high sensitivity and capacity of the 2-D matrix, it generates signatures of outputs of massively parallel DNA compute operations. These signatures can be fed to a silicon computer for analysis. On the surface this may seem redundant; why not just given the entire problem to a silicon computer from the start? The answer is that the 2-D matrix can reduce an exponential problem to a polynomial one. The DNA compute information obtained is based upon the recognition of patterns of hybridizations to the huge DNA code immobilized in the 2-D matrix. The 2-D matrix can't give the answer completely, but it can organize and sort massive amounts of data, preprocessing information and thus making it feasible for silicon computer to solve the problem.

### 3.0 METHODS, ASSUMPTIONS, PROCEDURES

#### 3.1 Encoding Strategy to Incorporate Modifications into DNA Strands

The plasmid DNA pBluescript was used as the starting hardware [8]. Labeled Polymerase Chain Reaction, PCR, product was derived from the amplification of the approximately 200 base pair multiple cloning site of this plasmid with combinations of four different modifications: Alexa Fluor-488, BODIPY-FL, biotin, and digoxigenin (DIG). The writing step corresponds to incorporation of the modified nucleotides or primers using PCR. (See Figure 1.)

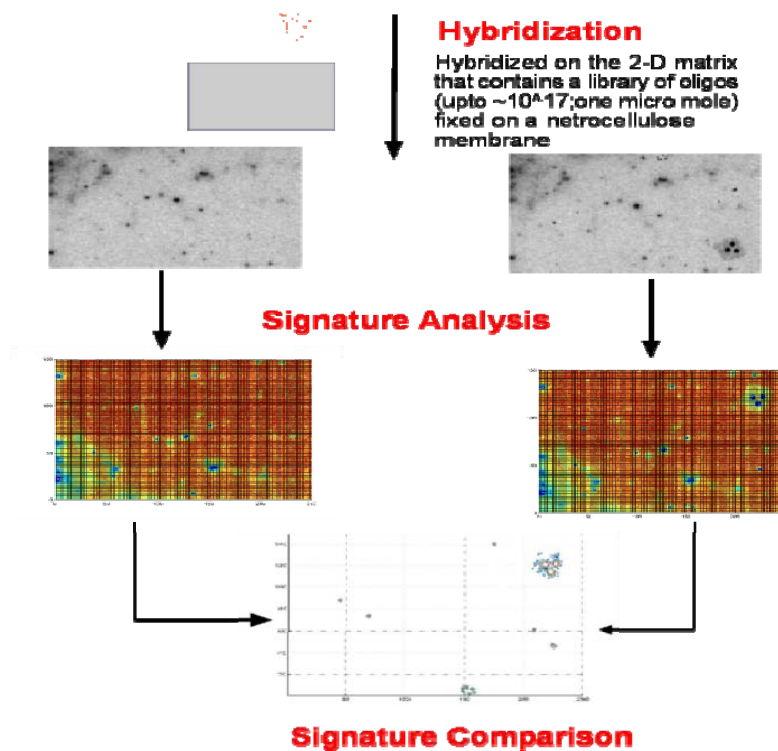


**Figure 1: Encoding Strategy**

In Figure 1, primers (P) and nucleotides (N) containing specific chemical modifications can be used in the PCR to create DNA fragments with those chemical modifications as shown. Specific chemical modifications include fluorescent compounds such as Alexa Fluor-488 and Bodipy- FL or non-fluorescent ones such as digoxigenin (DIG) and biotin. These four chemical modifications have successfully been incorporated into DNA fragments.

### 3.2 Coupling the Aqueous Method and the 2-D Array to Solve SAT

The 2-D DNA array can be used to solve the satisfiability (SAT) problem in the following way. Given the conjunction  $C_1 \wedge C_2 \wedge \dots \wedge C_k$  of disjunctive clauses  $C_i$  over  $n$  Boolean variables, the SAT problem is to decide if there is a Boolean string,  $x_1, x_2, \dots, x_n$ , for which the conjunction  $C_1 \wedge C_2 \wedge \dots \wedge C_k$  is true. This problem is known to be NP complete on a silicon computer. However, with  $k$ -copies of the universal 2-D array, SAT can be solved without having to use the many step mix and split method of [1]. After DNA discovery of the solution to a given clause  $C_i$  by the aqueous method (given in Section 4,) the DNA solutions to each clause can be dispersed on its own 2-D DNA array and the patterns formed on each array can be uploaded to a computer. The computer checks for common features of each pattern. This is mathematically equivalent to forming the intersection of the solution set to each clause. The intersection of the solutions sets to each clause  $C_i$  is the solution set to the conjunction of clauses,  $C_1 \wedge C_2 \wedge \dots \wedge C_k$ , i.e., a solution to SAT. Hence an exponential problem in  $n$  variables has been reduced to a polynomial one. The important nature of the 2-D array that makes this possible is that identical strands behave in the same manner anytime they are applied and different strands can be distinguished by their different behavior on the 2-D DNA array. (See Figure 2.)



**Figure 2: Signature Comparison for SAT Using Multiple 2-D Arrays**

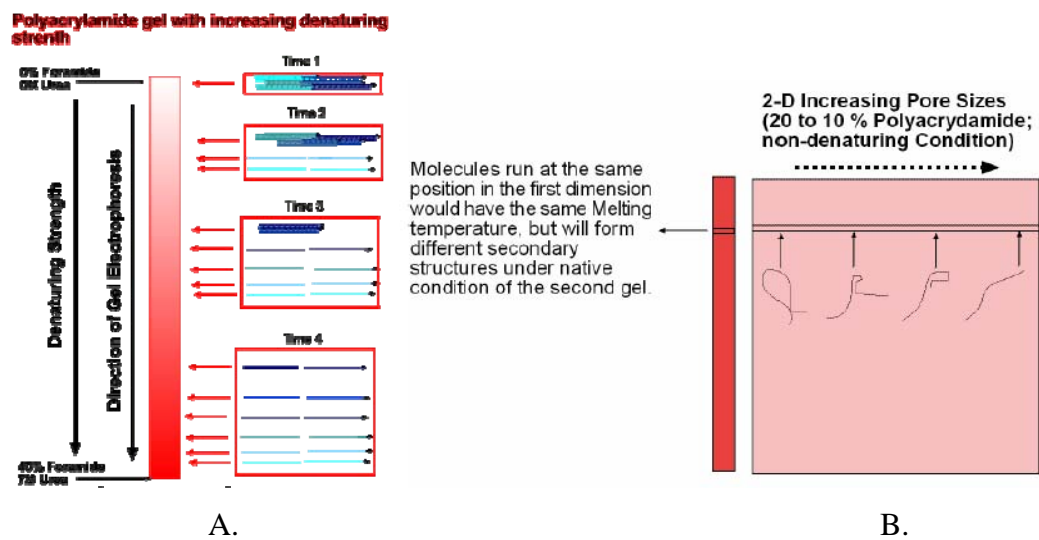
### 3.3 Construction of High Density 2-D array using SynDCode Generated DNA Codes

The DNA codes used below were generated by SynDCode [9]. The first step in building the 2-D matrix is to prepare the duplex DNA code to be used. A 70-nucleotide long single-stranded (ss) DNA that has 40 nucleotides of designed random sequences at its 5' end and a 20-mer long fixed sequence at the 3' end served as the primer site to convert this machine synthesized ssDNA-code to a dsDNA-code. The number of the different sequences in each code is limited by the DNA synthesizer machine. The current limitation for each synthesis by most DNA synthesizers is about one micro mole which is about  $10^{17}$  DNA molecules.

On this first dimension gel electrophoresis, the denaturing strength inside this polyacrylamide gel changes continuously from 0% to 100% (7 M urea plus 40% formamide) top to bottom. Once the electrophoresis started, the whole population duplex DNA molecules begin to encounter some low degree of denaturing strength (from urea, and formamide). Those dsDNA molecules that have lowest melting temperature start to denature, and those have higher melting temperature remain as duplex DNAs, and will be denatured later. Any duplex DNA molecule that has been completely denatured would become two single-stranded molecules, thus run faster than duplex DNAs under the denaturing condition. (See Figure 3A.)

Next, the strip of the first dimension gel with the denatured DNA is cut out, and laid on the top of a second gel. As shown in Figure 6, the DNAs in the first gel are now separated according to their melting temperatures, and those molecules that have roughly the same melting temperature are now migrated to about the same position in the first dimension of electrophoresis. The second gel is a native gel. Here, instead of denaturing the DNA; the secondary structures of the ssDNAs that have about the same melting temperature are maintained. In this second gel, the ssDNAs are separated according to their secondary structures formed by the intra-molecular hydrogen bonds. The DNA molecules will be further separated on the second gel resulting in a smear orthogonal to the original electrophoresis direction in the 2-D space on the second gel. This is the second dimension separation. (See Figure 3B.)

The second gel then is blotted on to a nitrocellulose membrane. This is 2-D matrix. The membrane can then be used as the universal array for detection of any input DNA. Figure 3A demonstrates the melting temperature dimension and Figure 3B demonstrates the secondary structure dimension.

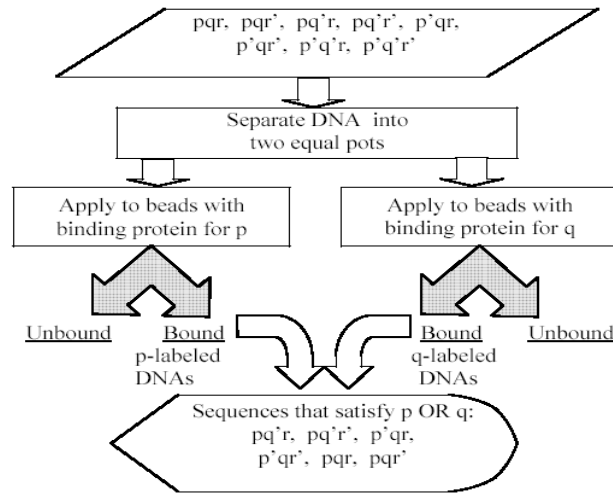


**Figure 3: Separation of DNA Code on 2-D array**

## 4.0 RESULTS & DISCUSSION

### 4.1 Mathematical Computation with Labeled Nucleotides

To implement the use of modified nucleotides into a logic problem, each modification represents a variable (e.g., “p”) that satisfies two conditions. The presence of the modification is taken as the true condition (p) while its absence is taken as the false condition (p’). For example, assign the DNA modified by Alexa Fluor-488 and biotin with the variables p and q, respectively. If a DNA molecule is labeled with the Alexa Fluor-488 modification, consider the molecule to be “True” at variable p. If it is labeled with biotin, consider the molecule to be “True” at variable q. The AND logical operator is satisfied when the DNA molecule contains all variables that satisfy the specified clause (e.g., DNA labeled with both biotin and Alexa Fluor-488 satisfies p AND q). The OR logical operator is satisfied when the DNA contains either one or both variables that satisfy the specified clause (e.g., DNA labeled with either biotin or Alexa Fluor-488 or both modifications satisfies p OR q).

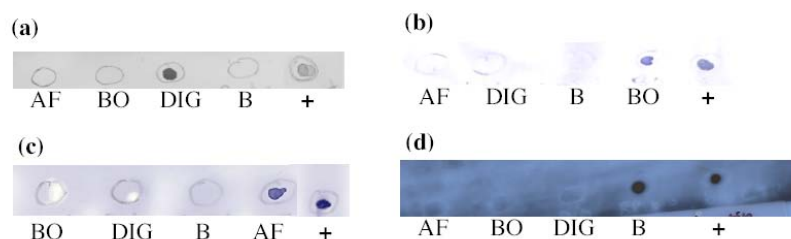


**Figure 4: Steps in Computing p OR q**

In Figure 4, labeled DNAs ('True') are represented by  $p$ ,  $q$  and  $r$ . Unlabeled DNAs ('False') are represented by  $p'$ ,  $q'$ ,  $r'$ . Each Boolean variable represents a distinct modification incorporated into the DNA. This figure represents a scheme for the isolation of the labeled DNAs (bound to the binding proteins) although this approach could alternatively be used for retention of unbound material.

Modified nucleotides are separated using specific binding proteins linked to magnetic beads to yield a solution to a SAT clause. (See Figure 4.) The specific modifications of interest bind to the beads (bound fraction) and are separated from other molecules using a small magnet. Unlabeled DNAs (unbound fraction) are removed from the tube, leaving only sequences that satisfy the specified clause. Thus, we can isolate either the molecules with the modification or those without it allowing one modification to represent both  $p$  and  $p'$ . Solution is confirmed through detection of the presence or absence of the modification on the DNA via dot blot [8]. (See Figure 5.)



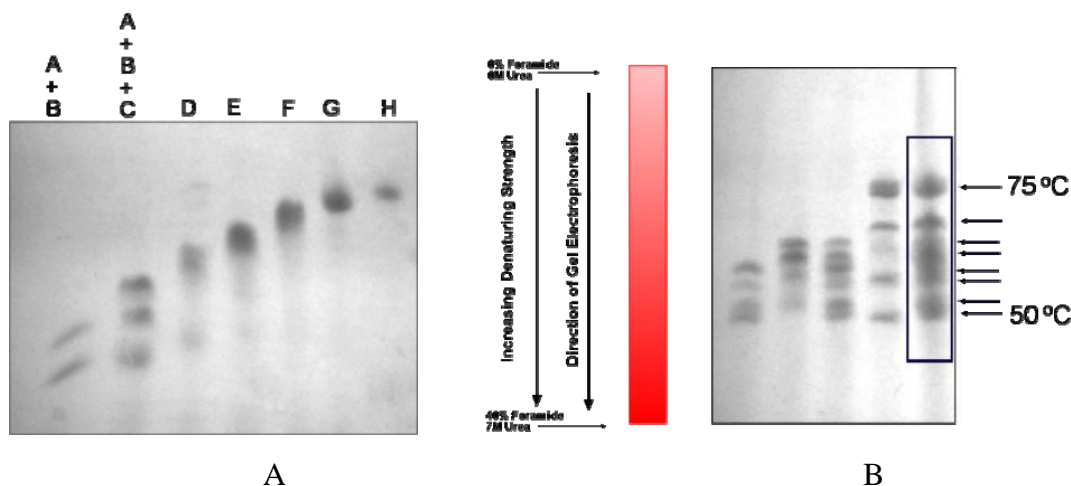


**Figure 5: Dot Blots Results**

In Figure 5, PCR products and controls labeled with specific modified nucleotides (AF = Alexa Fluor-488, BO = BODIPY-FL, DIG = digoxigenin, B = biotin, + = positive control) were dotted onto nylon membrane and treated with UV light to crosslink the DNA permanently to the membrane. Membranes were incubated with specific antibodies or binding proteins (a) anti-DIG, b) anti-BODIPY, c) anti- Alex Fluor, and d) streptavidin (binds biotin)) to identify the modified products. Binding proteins were then localized either using alkaline phosphatase (a, b, & c) or horse radish peroxidase as described in the methods. Dark spots on the membranes indicate the presence of antibodies or binding proteins at that location. Binding proteins recognized only specific modifications and not other unrelated compounds.

## 4.2 Using SynDCode to Test 2-D Array

Using a software upgrade of SynDCode [9] that designed strands to fit a desired DNA melting temperature gradient, we performed experiments to demonstrate that the denaturing gradient gel electrophoresis can separate duplex DNAs of same length according to their melting temperatures and this method is very precise and reproducible.



**Figure 6: Verification of Temperature Dimension Separation**

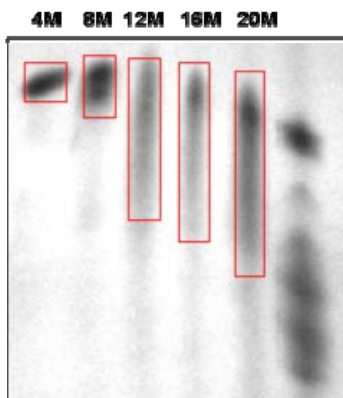
Initially demonstrated was that the denaturing gradient gel electrophoresis system separated eight designed 40-bp duplex DNA markers. The results indicate the duplex DNAs with the same length can be separated according to their melting temperatures. Repeated experiments indicate that the duplexes with lower melting temperatures always migrate fastest in the denaturing. The orders of the migrations of the duplexes were very reproducible. (See Figure 6A.)

To further characterize the denaturing gradient gel electrophoresis system, Figure 6B shows the separation of 15 SynDCode designed duplex DNAs according to their melting temperatures. The first dimension denaturing gradient gel system cannot separate those DNA molecules with similar melting temperatures, because the principle of the separation is based on the fact that each duplex DNA of same length should have a unique melting temperature and would be melted at its unique melting temperature. Therefore, in the last lane of Figure 6B, when we mixt all the 15 samples in one lane, we cannot detect 15 different bands on the gel, because several of them have almost identical melting temperatures as shown in Figure 7. Instead we can observe eight distinct bands representing the different groups of melting temperatures in the 15 samples (indicated by eight different colors in Figure 7). However each band has the second, secondary structure to distinguish sequences. This is why the second dimension is important.

SynDCode Strands	Melting Temp
ATTTTATATAATGTTAGAAATTTATTATTATCTTTAATAT	50.1°C
TTATTCAAAATATTTTCTATTATATTTAAACATATAAAAT	50.2°C
TCATATGTCATATTTATTATTGTGATTTATAAACATGAT	53.9°C
GATAATTTTGCCAGAGAAATAATATTCAAATAAAATGAAT	55.3°C
CTTAACACGTTATATTGGCGTATGTTTTATATGATCAA	57.9°C
CATCTATACCGTTAATGCCCAATACAATAGGACAATTTT	59.5°C
TTTAACATGATTCTTGTGTGGTTGTGTATATCGCATCTA	61.5°C
CATTGATACCCAGTTGGTTCTAATAATCATACTGAACGTC	60.6°C
AGTAGGTGAGATAGAGGAGAACTCAGACTCATGGTAAGCT	64.1°C
TGCATTCAATCTCGTGTCAATGGAGAAGACACGACTGGTG	66.2°C
CTCGTGTAGTGAGATTGCGTGCTCTATTGTTTTCAAGG	65.0°C
CTCAGGATCTTCATGCTACGCCCAGCCGATGTACGGTTGA	68.9°C
CTGGGTAAAGCAGCGTTCAAGTGGCAAGTGATCCGACCTC	68.7°C
CTTTAGGCCCGAAGTAGGGTGGCCAGTAGACCTACCGGC	71.0°C
GGCCCGGCTGCGGGTACTTGAACCTGACTCTTGCCGGGCC	75.4°C

**Figure 7: SynDCode Test Strands**

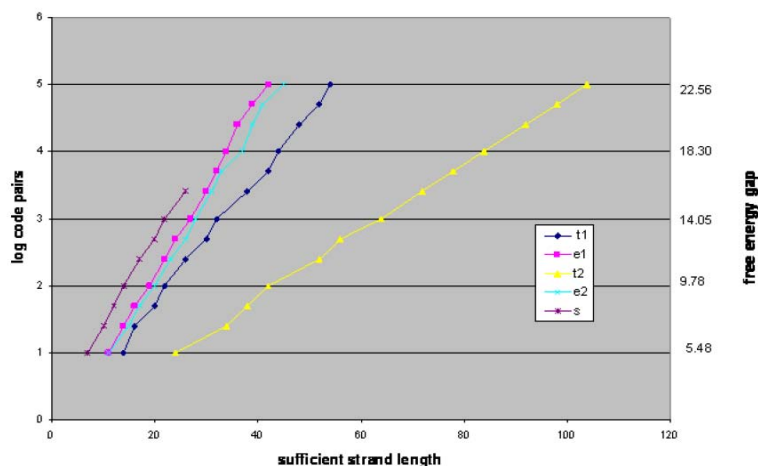
To test the gradient denaturing gel system in separating large number of different DNAs with different sequences, we used several SynDCode duplex DNA codes of sizes  $2^4$  to  $2^{20}$ . We ran these samples side by side in the gradient denaturing gel system. As expected, the results indicate as the number of different sequences in the code increased, the DNAs in the code are separated in an increasing area and quite evenly in the gel, indicated by a long smear marked by the red rectangles (See Figure 8.)



**Figure 8: Large Code Dispersion**

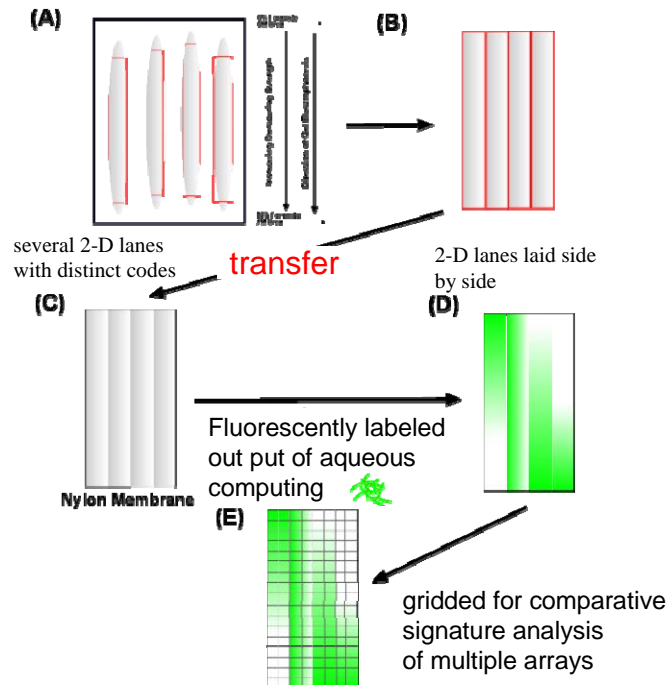
### 4.3 Obtaining a Dense 2-D Array

Figure 3 diagrams the idea of the 2-D matrix and Figure 7 diagrams the process of making it more dense and applying it to the output of the modified nucleotide aqueous computing system. Our experiments have shown that the DNA in this system behaves consistently and reproducibly. Using eight distinct 50-mer DNA codes, a total separation of  $2^{21}$  50-mers on a nylon membrane was achieved. Figure 9 and ref. [10] show that error-free 50-mer codes of size  $2^{21}$  exist and that SynDCode can find them. In Figure 9, extrapolation of line e1 indicates that 50-mers are sufficient for a code of size  $2^{21}$ .



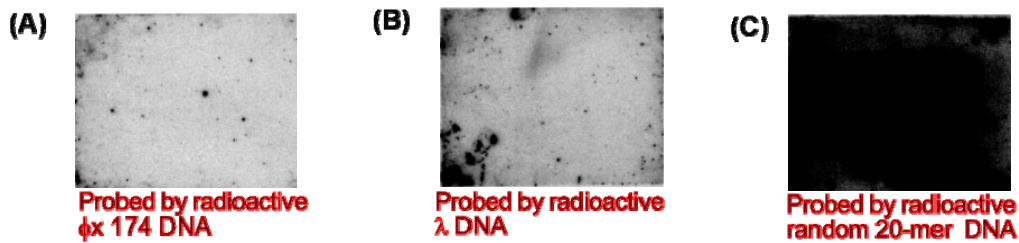
**Figure 9: Sufficient Strand Length for Error-Free Codes of Given Sizes**

To make the dense 2-D matrix, eight distinct DNA codes of size  $2^{21}$  were separated in the denaturing gradient gel system as shown in Figure 8. After staining the gel (with ethidium bromide or other reagents), we then cut the area containing the DNAs from the gel. Next, the cut out gel strips are aligned and transferred to a commercially available membrane such as super charged nylon from Schleicher& Schuell by standard blotting methods (indicated by the red box in Figure 10A). The resulting nylon membrane now has the DNA codes attached on the surface can be used to probe any radioactive or fluorescent labeled input DNA.



**Figure 10: Making and Using a Dense Array**

To verify the separation, this nylon membrane was probed with 3 different samples having increasing complexity. In the first experiment, the 2-D matrix was probed by radioactive probes using  $\phi$ X 174 DNA (~ 4 kb) as template. As shown in Figure 11A, several spots in the 2-D matrix indicate hybridization. After stripping off the first sample, the same 2-D matrix was probed by probes made from  $\lambda$  DNA (~40 kb). It is clear that in this case more areas of the 2-D matrix show hybridization (Figure 11B). To show that the DNA molecules from the 8 libraries are spread in the nylon membrane, we probed the membrane with radioactive labeled all possible 20-mers ( $4^{20}$  different sequences). As shown in Figure 11C, the radioactive 20-mer random sequences hybridized to nearly the entire 2-D matrix space indicating that the 2-D matrix indeed contains duplex DNA molecules covering the whole 2-D space.



**Figure 11: Verification of Dense Array Construction**

## 5.0 CONCLUSION

In regard to labeled nucleotides to perform computations, the experiments demonstrate that DNA can be successfully modified to contain a variety of fluorescent and non-fluorescent labels and that these modifications can be used to do computation. The results presented here allow for the creation of a four variable library with labeling of only four, as opposed to eight, sites. We expect to be able to scale-up this approach since there are at least 12 fluorescent and 3 non-fluorescent modifications with commercially available binding proteins. With all of these chemical compounds incorporated, either as modified primers or nucleotides, we would be able to solve 15-variable problems. Our work shows that when as more labels become available, more computing power can be obtained.

In regard to the DNA array construction, a DNA code of total size  $10^7$  50-mers were distributed on a two dimensional array. This exceeds the current  $10^6$  capacity of photolithographic or ink-jet arrays and will allow for the solution of a 21 variable SAT problem. This array could potentially have a  $10^{17}$  capacity, allowing for the solution of a 50 variable SAT.

Different sequences end up at different locations in this 2-D array. The distribution of different sequences is self-directed, systematic and continuous. The 2-D matrix is universal in its application and easy to make. A collection of fixed sequences can be added to the 2-D matrix to generate a grid inside the 2-D matrix for enhanced applications, however the 2-D matrix is self calibrating, because it seeks to achieve recognizable global patterns and signal-processing tools can extract information.

## **5.1 Other Applications**

One bio-defense application is that the 2-D genomic matrix system proposed here is capable of capturing a snapshot of an entire biosystem's state. At a later time, the snapshot can be updated, or a new one acquired, and compared to a previous memory to measure changes such as the appearance of pathogens. For example, this tool could be used for monitoring for bioterrorism agents in water supplies. The goal is to use genomic information at the population or community scale to monitor and detect the existence of new biota (such as pathogens) in the environment. The scope of organisms with their genomic DNA sequenced is fairly small. Thus, much information at the genomic level is not available with conventional techniques. The 2-D genomic matrix provides a way to access information from all organisms in a community and a way to assess impact by human and non-human biomaterials. The 2-D genomic matrix does not require explicit sequence knowledge and it is quick, flexible, and inexpensive to implement. Thus, it could provide a better diagnostic tool for environmental monitoring that provides a holistic view of the genomic status of an ecosystem.



## 6.0 REFERENCES

1. Adleman, L. M., "Molecular computation of solutions to combinatorial problems," *Science*, **266**, 1994 pp. 1021–1024.
2. Head, T. and Gal, S. "Aqueous computing: writing into fluid memory," *Bulletin of the European Association for Theoretical Computer Science*, **75**, 2001 pp. 190-198.
3. Frutos, A. G. et al. "Demonstration of a word design strategy for DNA computing on surfaces," *Nucleic Acids Research*, **25**, 1997 pp. 4748 -4756.
4. Murphy, D., "Gene expression studies using microarrays: principles, problems, and prospects," *Advances in Physiology Education*, **26**, 2002 pp. 256–270.
5. Winfree, E., et al. "Design and self-assembly of two-dimensional DNA crystals," *Nature*, **394**, 1998 pp. 539–544.
6. Braun, E., et al. "DNA-templated assembly and electrode attachment of a conducting silver wire," *Nature*, **391**, 1998 pp. 775–778.
7. Whitesides G. M. and Boncheva, M., "Beyond molecules: Self-assembly of mesoscopic and macroscopic components," *Proc. Natl. Acad. Sci.*, **99**, 2002 pp. 4769–4774.
8. Pagano A. M and Gal, S. "An approach for using modified nucleotides in aqueous DNA computing," *Proceedings DNA 14*, 2007, to appear
9. M. A. Bishop, A. J. Macula, T. E. Renz, SynDCode Suite, <http://syndcode.geneseo.edu>.
10. M. A. Bishop, et al. "Free energy gap and statistical thermodynamic fidelity of DNA codes," *J. Comp. Biol.* **14**, 2007 pp. 1088-1104.